ELECTRONICALLY FILED

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of:	Barbara Ensoli	Confirmation No :	9400
Application No :	09/555,534	Art Unit:	1648
Filed:	May 31, 2000	Examiner:	Humphrey, Louise Wang Zhiying
For:	HIV TAT, OR DERIVATIVES THEREOF FOR PROPHYLACTIC AND THERAPEUTIC VACCINATION	Attorney Docket No :	11340-003-999

THIRD DECLARATION OF MAURO MAGNANI, Ph.D.
UNDER 37 C.F.R. § 1.132

Mail Stop RCE
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

I, MAURO MAGNANI, Ph.D., do declare as follows:

1. I am currently Professor of Biochemistry, Director of Centre of Biotechnology, Vice-Rector of the University of Urbino, Urbino, Italy. I have over thirty years of experience as a biochemist in the research and development of products and applications useful in the biotechnology and pharmaceutical industries. I am included in the official list of professional biologists in Italy with n. 017484 "Ordine Nazionale Biologi," and I am a Technical Director nominated by the "Agenzia Italiana del Farmaco, AIFA" with n. AIDT-19/2005. My education and experience are summarized on my Curriculum Vitae, which is attached hereto as Exhibit 1.

2. I have collaborated with Dr. Barbara Ensoli, who is the inventor of the above-identified application No. 09/555,534 (hereinafter "the '534 application"), in the development of methods to produce recombinant, biologically active Tat protein. I also supervise and have supervised the production of such biologically active Tat protein, according to good manufacturing practices (GMP), for use in human clinical trials. In particular, I have

experience in reversed-phase high pressure liquid chromatography (RP-HPLC) purification of biologically active Tat protein, which my laboratory has carried out many times

3. I have been asked to review Gu *et al.* Sep Technol , 1994, 4:258-260 ("Gu *et al.*"), and to consider, based on my experience, whether the phase separation method described in Gu *et al* is applicable to removing acetonitrile from compositions comprising biologically active Tat protein purified by RP-HPLC For the reason discussed below, it is my judgement and opinion that the phase separation method of Gu *et al.* is not applicable to removing acetonitrile from compositions comprising biologically active Tat protein purified by RP-HPLC.

4. Gu *et al.* discloses removing acetonitrile from RP-HPLC effluent fractions by a phase separation method (see Abstract) In particular, Gu *et al* discloses that when a RP-HPLC effluent fraction containing 65% (vol) acetonitrile/35% water/0 1% trifluoroacetic acid (TFA) is stored in a freezer at -17°C for several hours, a phase separation occurs such that a top phase containing 88% (vol.) acetonitrile, and a bottom phase containing 65% (vol) water and 99%+ of the human growth hormone (hGH) protein, are formed (see Abstract; and page 258, right column, first paragraph) Gu *et al.* indicates that the phase separation "occurs *only* in the [acetonitrile] concentration range of 35-88%" (see page 259, right column, second paragraph, lines 3-4) (emphasis in original), and explains that the reason the hGH protein and its genetically engineered analog (hGHG120R) stay in the bottom phase is probably due to their hydrophilicity (see page 260, sentence bridging left and right columns)

5 In my experience, Tat proteins elute during RP-HPLC at acetonitrile concentrations between 25 and 30%, more precisely at acetonitrile concentrations between 28 and 30% I have never observed any Tat protein to elute during RP-HPLC at an acetonitrile concentration higher than 35% Therefore, it is my judgment and opinion that the phase separation method of Gu *et al* is not applicable to removing acetonitrile from compositions comprising biologically active Tat protein purified by RP-HPLC, since the Tat protein does not elute at the required acetonitrile concentration range of 35-88% during RP-HPLC.

6 Attached as Exhibits 2-4 are hydrophilicity profiles of the Tat protein, hGH protein, and hGHG120R analog, respectively I have been informed that these profiles were obtained using ProtScale (http://www expasy org/tools/protscale html), which allows

computation and representation (in the form of a two-dimensional plot) of the hydrophilicity profile of selected proteins based on the amino acid scale of Hopp and Woods, "Prediction of protein antigenic determinants from amino acid sequences," Proc Natl Acad Sci U S A, June 1981, Vol 78, No 6, pp 3824-3828 ("Hopp & Woods"), a copy of which reference is attached as Exhibit 5 I am informed that the sequence of the Tat protein used to generate its hydrophilicity profile was the sequence disclosed in the '534 application at page 37, lines 9-10 I am also informed that the sequence of the hGH protein used to generate its hydrophilicity profile was the amino acid sequence having GenBank accession number AAA72260, obtained on December 17, 2008 from the online database of the National Center for Biotechnology Information (NCBI) at http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&id=208528, which is 192 amino acids in length (see printout of the record for GenBank accession number AAA72260, which is attached hereto as Exhibit 6) or one amino acid residue longer in length than that disclosed in Gu *et al.* at page 258, left column, second paragraph, lines 1-2, and that the sequence of the hGHG120R analog used to generate its hydrophilicity profile is the sequence of GenBank accession number AAA72260, except that the glycine (G) residue at position 121 is replaced with arginine (R) Briefly, according to the method of Hopp & Woods, to generate the hydrophilicity profile of a protein, each amino acid residue in the sequence of the protein is assigned its hydrophilicity value, then these values are repetitively averaged down the length of the polypeptide chain, generating a series of local hydrophilicity values, which are averaged at each repetition in groups of six (see Hopp & Woods, page 3824, right column, fourth paragraph, lines 1-7) As evidenced by the hydrophilicity profiles in Exhibits 2-4, the profile for the Tat protein differs markedly from the profiles for the hGH protein and hGHG120R analog, respectively, in that the profiles for the hGH protein and hGHG120R analog contain many more high points of local hydrophilicity, dispersed along the length of the proteins, than are found in the profile for the Tat protein It is my judgment and opinion that, based on the differences in the hydrophilicity profiles between on the one hand, the hGH protein and hGHG120R analog, and on the other hand, the Tat protein, one of ordinary skill in the art would not expect that biologically active Tat protein would be preferentially present in the bottom, predominantly water, phase of Gu *et al.*'s phase separation method; rather, one would expect that much of the Tat protein would be removed along with the acetonitrile phase or remain at the interface between the water phase and acetonitrile phase.

7. I also have been asked to review U.S. Patent No. 5,646,120 by Sumner-Smith *et al.* ("Sumner-Smith *et al.*") and to consider, based on my experience, whether an acid exchange reaction such as that described by Sumner-Smith *et al.* could be used to exchange the TFA present in a composition with biologically active Tat protein after RP-HPLC, without loss of biological activity of the Tat protein. For the reasons discussed below, it is my judgement and opinion that an acid exchange reaction such as that described by Sumner-Smith *et al.* would be expected to destroy the biological activity of Tat protein purified by RP-HPLC.

8. In my experience, a Tat protein is completely denatured during RP-HPLC that occurs in a non-aqueous acetonitrile/TFA solvent. If the Tat protein is lyophilized after RP-HPLC, most of the acetonitrile and TFA will be removed by the lyophilization, since acetonitrile and TFA are volatile; if the Tat protein is then suspended in a buffer compatible with the biological activity of the Tat protein (*i.e.*, in an aqueous, pH-buffered, neutral solvent), in this compatible solvent, the Tat protein regains its native conformation and biological activity. However, if the biologically active Tat protein and remaining TFA after RP-HPLC elution, lyophilization, and resuspension is exchanged with another acid in an aqueous solvent, such as by being subject to the acid exchange reaction disclosed in Sumner-Smith *et al.* at col. 9, lines 44-62, I would expect the three-dimensional conformation of the Tat protein to be damaged by the acid, which acid is chemically reactive in the aqueous solvent, and thus the Tat protein would be expected to lose biological activity. Thus, in my judgement and opinion, the acid exchange reaction described by Sumner-Smith *et al.* would destroy the biological activity of HIV Tat protein purified by RP-HPLC.

9. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that I make these statements with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of this application, and any patent issuing thereon.

Date: January 3th, 2009

Mauro Magnani
Professor of Biochemistry

EXHIBIT 1

Curriculum Vitae – Prof. Mauro Magnani

MAGNANI Prof. Mauro, Ph.D. Italian, Professor of Biochemistry
BORN: April 9, 1953, Italy
LANGUAGES: Italian, English
EDUCATION: Univ. Urbino, Italy, Ph.D., 1976
PRIMARY POSITION: Professor of Biochemistry and Director Centre of Biotechnology.
PROFESSIONAL CAREER: Visiting Researcher, Dept. Biochemistry, Univ. Birmingham, 1980; Visiting Prof. Dept. Biolgy, Haifa, Israel, 1983; Asst. Prof. Univ. Urbino, 1977-82, Assoc. Prof., 1982-1986, Prof. 1986 - ; Dean, Faculty of Sciences University of Urbino 1995-2001; Director Interuniversity Consortium for Biotechnology (CIB) 1998-2004; Vice Rector of the University of Urbino 2001- .Include in the official list of professional biologist in Italy with n. 017484 "Ordine Nazionale Biologi". Technical Director nominated by the "Agenzia Italiana del Farmaco, AIFA" with n. AIDT-19/2005.
CURRENT RESEARCH: Development of new drug delivery and drug targeting systems; Protein turnover ubiquitination and regulation of gene expression; Mechanisms of drug resistance and drug toxicity; Modulation of NF-kB and gene expression by oligonucleotide decoys, vaccine development; nanobiotechnology in drug delivery.
PUBLICATIONS: over 300 articles published in international refereed scientific journals; Co-editor of three books:
"*Red Blood Cell Aging*", Plenum Press, N.Y., 1991, pp. 383.
"*The Use of Resealed Erythrocytes as Carriers and Bioreactors*", Plenum Press, N.Y., 1992, pp. 361.
"*Erythrocyte Engineering for Drug Delivery and Targeting*", Landes Bioscience, 2002.
REFEREE: Programmes of the E.U.; The International Science Foundation (U.S.A.); Target Project "Biotechnology" of the National Research Council (C.N.R.).; Member of the Project "Patologia clinica e terapia dell'infezione da HIV" of the Italian Ministry of Health; PRIN and FIRB Projects of Italian Ministry of University and Research; Member of Committee Post Genoma (C.N.R); Include in the "Albo degli Esperti" of M.I.U.R. and Eureka Projects of EU.
REVIEWER: Biotechnology and Applied Biochemistry; Nature Biotechnology; Drugs; Leukemia; European Journal Haematology; Biochimica et Biophysica Acta; Blood; Journal of Cellular Engineering; Journal of Internal Medicine; Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology; Mechanisms of Ageing and Development; Antiviral Research; Journal of Chromatography; Journal of Biological Regulators and Homeostatic Agents; Life Sciences; Biochemistry; International Journal of Biochemistry and Cell Biology; Human Gene Therapy; European Journal of Biochemistry; Clinical Pharmacokinetics; Autoimmunity; Oncogene; Haematologica; J. Controlled Release; Editorial Board: Current Drug Targets, Biotechnology.

PATENTS

Europen Patent EP 0517986B1
M. Magnani, L. Rossi "*Transformed erythrocytes, process for preparing the same, and their use in pharmaceutical compositions*"

US Patent 5,753,221

M. Magnani, L. Rossi "*Transformed erythrocytes, process for preparing the same, and their use in pharmaceutical compositions*"

US Patent N. 6.139.836
Mauro Magnani, Ivo Panzani, Leonardo Bigi, Andrea Zanella "*Method of encapsulating biologically active agents within erythrocytes, and apparatus therefor*".
Assignee: Dideco S.p.A., Mirandola, Italy

European Patent N. EP98830479.6
M. Magnani, G. Brandi, A. Fraternale, A. Casabianca "*Pharmaceutical composition or composition package containing a pyrimidine nucleoside analogue and a purine nucleoside analogue*".

Brevetto C.N.R. N. RM92 A 000377
M. Magnani "*Antigeni legati alla superficie esterna di eritrociti e procedimento per la loro preparazione*"

Brevetto C.N.R. N. RM 93 A 000474
M. Magnani "*Eritrociti incorporanti alcool ossidasi e loro uso nelle intossicazioni da metanolo*"

Brevetto C.N.R.
M. Magnani, L. Rossi, G. Brandi, E. Millo, G. Damonte, U. Benatti, A. De Flora "*Profarmaco di acyclovir e suo uso in composizioni farmaceutiche*"

Brevetto di Invenzione N. MI2002A01196 – 06/06/1996 – PCT/IT 02/00368 del 13/06/2002
M. Magnani, C. Fiorucci, P. Filippone, G. Brandi, M. Paiardini. "*Derivato tetramerico dell'indol-3 carbinolo ad attività anticancerogena e metodo di sintesi del derivato stesso*".

Brevetto di Invenzione N. TO2001A01077 – 16/11/2001
M. Magnani, F. Graziano, A. Ruzzo "*Mutazioni della linea germinale nel promotore del gene della E-caderina e metodi di diagnosi per individuare una maggiore suscettibilità al carcinoma gastrico*".

Brevetto N. TO2003A001048 – 30/12/2003 - PCT/EP/2004/053726 – 29/12/2004
U.Benatti, G. Brandi, E. Garaci, M. Magnani, E. Millo, A.T. Palamara, L. Rossi. "*Derivati del glutathione e loro utilizzo per il trattamento di malattie virali*".

EXHIBIT 2

ProtScale

User-provided sequence:

```
        10         20         30         40         50         60
MEPVDPRLEP WKHPGSQPKT ACTNCYCKKC CFHCQVCFIT KALGISYGRK KRRQRRRPPQ
        70         80
GSQTHQVSLS KQPTSQSRGD PTGPKE
```

SEQUENCE LENGTH: 86

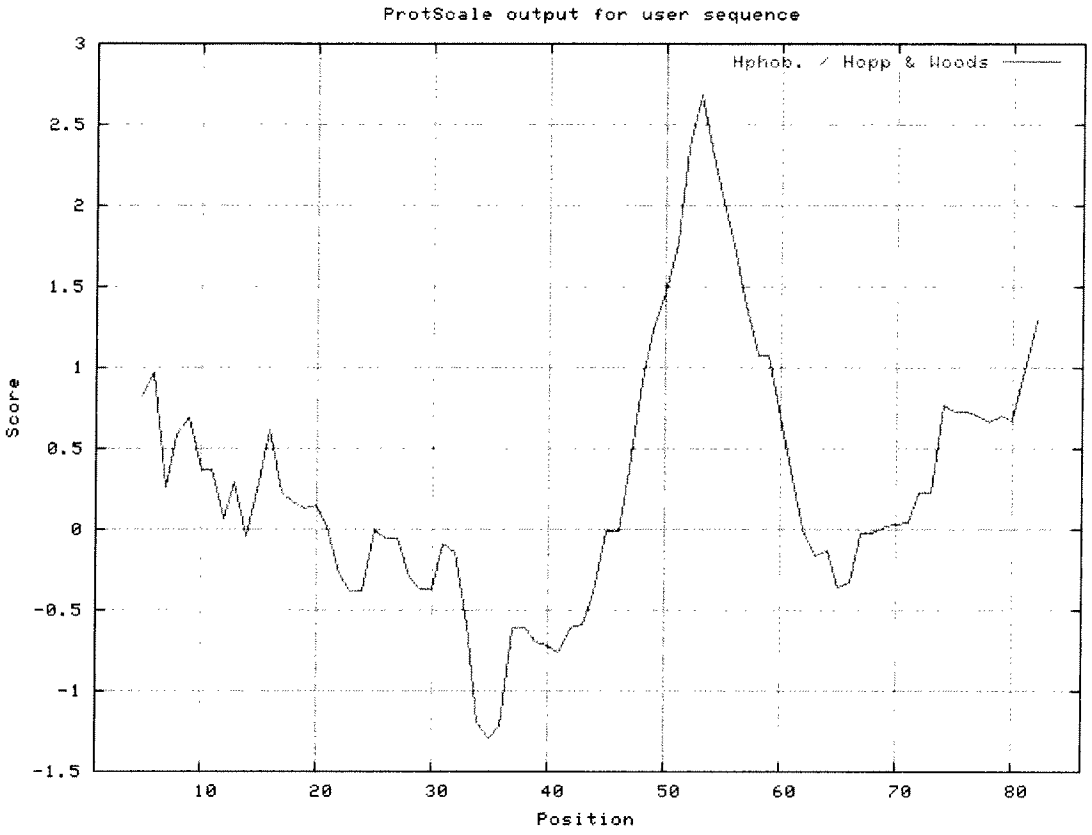
Using the scale **Hphob. / Hopp & Woods**, the individual values for the 20 amino acids are:

```
Ala: -0.500  Arg:  3.000  Asn:  0.200  Asp:  3.000  Cys: -1.000  Gln:  0.200
Glu:  3.000  Gly:  0.000  His: -0.500  Ile: -1.800  Leu: -1.800  Lys:  3.000
Met: -1.300  Phe: -2.500  Pro:  0.000  Ser:  0.300  Thr: -0.400  Trp: -3.400
Tyr: -2.300  Val: -1.500  :  1.600  :  1.600  : -0.215
```

Weights for window positions 1,..,9, using **linear weight variation model**:

```
   1     2     3     4     5     6     7     8     9
 1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
 edge                    center                  edge
```

```
MIN: -1.289
MAX: 2.689
```



The results of your ProtScale query are available in the following formats:

- Image in GIF-format
- Image in Postscript-format
- Numerical format (verbose)
- Numerical format (minimal, to be exported into an external application)

EXHIBIT 3

ProtScale

User-provided sequence:

```
        10         20         30         40         50         60
MFPTIPLSRL FDNAMLRAHR LHQLAFDTYQ EFEEAYIPKE QKYSFLQNPQ TSLCFSESIP
        70         80         90        100        110        120
TPSNREETQQ KSNLELLRIS LLLIQSWLEP VQFLRSVFAN SLVYGASDSN VYDLLKDLEE
       130        140        150        160        170        180
GIQTLMGRLE DGSPRTGQIF KQTYSKFDTN SHNDDALLKN YGLLYCFRKD MDKVETFLRI
       190
VQCRSVEGSC GF
```

SEQUENCE LENGTH: 192

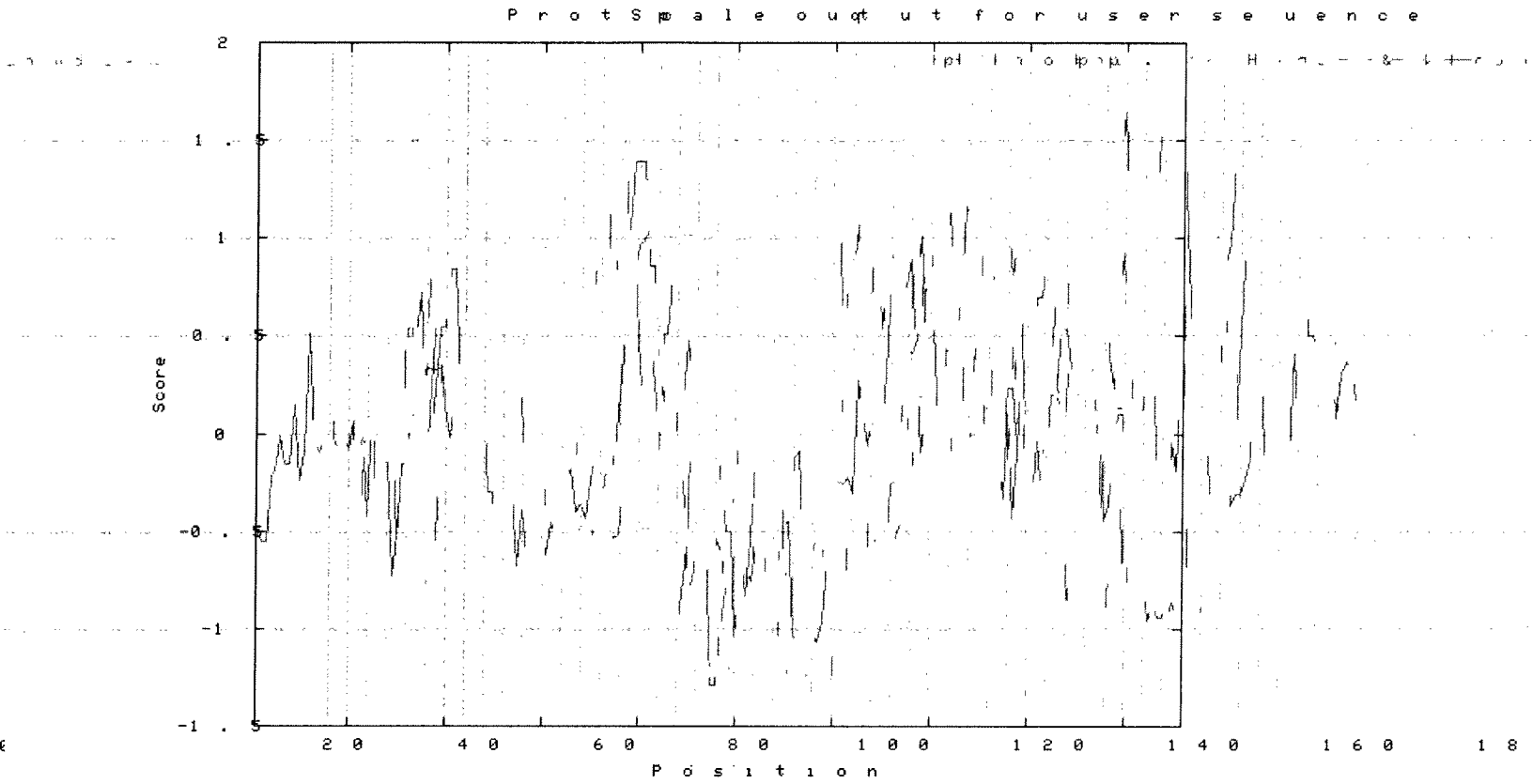
Using the scale **Hphob. / Hopp & Woods**, the individual values for the 20 amino acids are:

```
Ala: -0.500  Arg:  3.000  Asn:  0.200  Asp:  3.000  Cys: -1.000  Gln:  0.200
Glu:  3.000  Gly:  0.000  His: -0.500  Ile: -1.800  Leu: -1.800  Lys:  3.000
Met: -1.300  Phe: -2.500  Pro:  0.000  Ser:  0.300  Thr: -0.400  Trp: -3.400
Tyr: -2.300  Val: -1.500  :  1.600  :  1.600  : -0.215
```

Weights for window positions 1,..,9, using **linear weight variation model**:

```
   1     2     3     4     5     6     7     8     9
 1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
 edge                    center                  edge
```

```
MIN: -1.289
MAX: 1.644
```



The results of your ProtScale usage are available in the following formats:

- Image in GIF format
- Image in Postscript-format
- Numerical format (verbose)
- Numerical format (minimal, to be exported into an external application)

EXHIBIT 4

ProtScale

User-provided sequence:

```
        10         20         30         40         50         60
MFPTIPLSRL FDNAMLRAHR LHQLAFDTYQ EFEEAYIPKE QKYSFLQNPQ TSLCFSESIP
        70         80         90        100        110        120
TPSNREETQQ KSNLELLRIS LLLIQSWLEP VQFLRSVFAN SLVYGASDSN VYDLLKDLEE
       130        140        150        160        170        180
RIQTLMGRLE DGSPRTGQIF KQTYSKFDTN SHNDDALLKN YGLLYCFRKD MDKVETFLRI
       190
VQCRSVEGSC GF
```

SEQUENCE LENGTH: 192

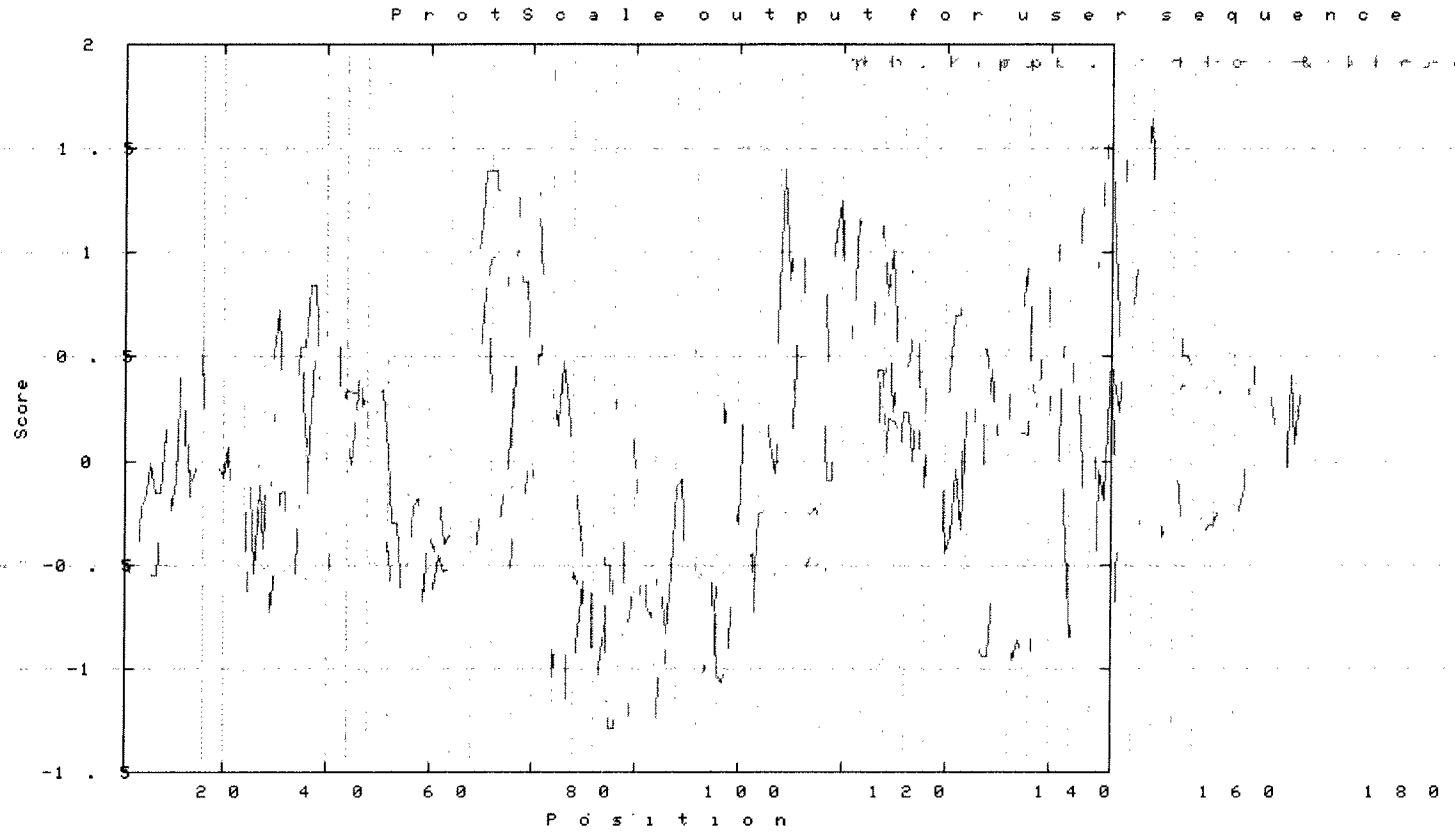
Using the scale **Hphob. / Hopp & Woods**, the individual values for the 20 amino acids are:

```
Ala: -0.500  Arg:  3.000  Asn:  0.200  Asp:  3.000  Cys: -1.000  Gln:  0.200
Glu:  3.000  Gly:  0.000  His: -0.500  Ile: -1.800  Leu: -1.800  Lys:  3.000
Met: -1.300  Phe: -2.500  Pro:  0.000  Ser:  0.300  Thr: -0.400  Trp: -3.400
Tyr: -2.300  Val: -1.500  :  1.600  :  1.600  : -0.215
```

Weights for window positions 1,..,9, using **linear weight variation model**:

```
   1     2     3     4     5     6     7     8     9
 1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
 edge                    center                  edge
```

```
MIN: -1.289
MAX: 1.644
```



The results of your ProtScale query are available in the following formats:

- Image in GIF-format
- Image in Postscript-format
- Numerical format (verbose)
- Numerical format (minimal, to be exported into an external application)

EXHIBIT 5

Proc. Natl. Acad. Sci. USA
Vol. 78, No. 6, pp. 3824–3828, June 1981
Immunology

Prediction of protein antigenic determinants from amino acid sequences

(hydrophilicity analysis/protein conformation)

THOMAS P. HOPP AND KENNETH R. WOODS

The Lindsley F. Kimball Research Institute, The New York Blood Center, 310 East 67th Street, New York, New York 10021

Communicated by Bruce Merrifield, March 2, 1981

ABSTRACT **A method is presented for locating protein antigenic determinants by analyzing amino acid sequences in order to find the point of greatest local hydrophilicity. This is accomplished by assigning each amino acid a numerical value (hydrophilicity value) and then repetitively averaging these values along the peptide chain. The point of highest local average hydrophilicity is invariably located in, or immediately adjacent to, an antigenic determinant. It was found that the prediction success rate depended on averaging group length, with hexapeptide averages yielding optimal results. The method was developed using 12 proteins for which extensive immunochemical analysis has been carried out and subsequently was used to predict antigenic determinants for the following proteins: hepatitis B surface antigen, influenza hemagglutinins, fowl plague virus hemagglutinin, human histocompatibility antigen HLA-B7, human interferons, *Escherichia coli* and cholera enterotoxins, ragweed allergens Ra3 and Ra5, and streptococcal M protein. The hepatitis B surface antigen sequence was synthesized by chemical means and was shown to have antigenic activity by radioimmunoassay.**

The elucidation of protein antigenic structures is presently a difficult, uncertain, and time-consuming task. To precisely delineate antigenic determinants, it is necessary to prepare a large number of well-characterized chemical derivatives and peptide fragments from the original protein antigen and then to test these derivatives for immunological activity (1, 2). Alternatively, a homologous series of proteins may be used to assess the influence of particular amino acid substitutions, thereby implicating certain regions as antigenic determinants (3, 4); this approach requires knowledge of complete primary structures for a number of proteins before the immunological results can be interpreted. Despite the laboriousness of available approaches, the complete antigenic structures have been elucidated for a small number of proteins, and partial information is available for many others.

As more information becomes available on protein antigens, it should be possible to use this information to predict the locations of antigenic determinants before any immunological testing has been carried out. In recent years a number of systems have been developed to predict protein conformational features from amino acid sequences (5–8), but none of these were specifically oriented to the prediction of antigenic determinants. Therefore, we sought a method that was not predicated upon predictions of particular structural features but rather sought a simple correlation with surface location of stretches of peptide chain and the likelihood of antibody binding. A guiding principle was the notion that many surface oriented regions are *nonantigenic* (1). This led us to take an empirical approach in our analysis and to arbitrarily manipulate the emphasis placed on certain amino acids in order to find a particular kind of sequence that is favored for antibody binding (which may not strictly depend on the hydrophilicity of the sequence). The present report describes a system that uses a simplified method to successfully predict antigenic determinants, given the amino acid sequence of a protein and no other information.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

METHOD

Previous investigations have demonstrated that antigenic determinants are surface features of proteins and indicate that they are frequently found on regions of a molecule that have an unusually high degree of exposure to solvent—i.e., regions which project into the medium (for reviews, see refs. 1 and 3). This, together with the fact that charged, hydrophilic amino acid side chains are common features of antigenic determinants, led us to investigate the possibility that at least some antigenic determinants might be associated with stretches of amino acid sequence that contain a large number of charged and polar residues and are lacking in large hydrophobic residues. A suitable means of methodically searching for such regions was found by combining a method like that of Chou and Fasman (5), in which numerical values for amino acids are repetitively averaged over the length of a polypeptide chain, with a set of values expressing the relative hydrophilicity of each amino acid. Suitable values were available in the solvent parameters assigned by Levitt (6), which are derivatives of the hydrophobicity values of Nozaki and Tanford (9).

In Table 1 are listed the numerical values (hydrophilicity values) assigned to the 20 amino acids commonly found in proteins. In the first column, the values of Levitt (6) are listed, whereas the second column lists the values that were finally chosen for our hydrophilicity calculations. The values were generally retained as expressed by Levitt; however, changes in the values for proline, asparatic acid, and glutamic acid improve the prediction results, as explained later. Hydrophilicity analysis of a protein is carried out by the following method.

Each amino acid in the sequence of the protein is assigned its hydrophilicity value, then these values are repetitively averaged down the length of the polypeptide chain, generating a series of local hydrophilicity values. The number of hydrophilicity values that are averaged at each repetition is arbitrary, and we chose groups of six for our initial studies because this is the approximate size of an antigenic determinant (1, 10). Once the complete set of averaged values is obtained, the list is scanned to locate the highest value. According to the studies presented here, this high point will invariably lie within or be immediately adjacent to one of that protein's antigenic determinants.

A useful way of recording the results of this analysis is to produce a plot of hydrophilicity value versus sequence position.

Abbreviations: αAbu, α-aminobutyric acid; HBsAg, hepatitis B surface antigen.

Table 1. Hydrophilicity values

Amino acid	s,* kcal/mol	Hydrophilicity value
Arginine	3.0	3.0
Aspartic acid	2.5	3.0
Glutamic acid	2.5	3.0
Lysine	3.0	3.0
Serine	0.3	0.3
Asparagine	0.2	0.2
Glutamine	0.2	0.2
Glycine	0.0	0.0
Proline	−1.4	0.0
Threonine	−0.4	−0.4
Alanine	−0.5	−0.5
Histidine	−0.5	−0.5
Cysteine	−1.0	−1.0
Methionine	−1.3	−1.3
Valine	−1.5	−1.5
Isoleucine	−1.8	−1.8
Leucine	−1.8	−1.8
Tyrosine	−2.3	−2.3
Phenylalanine	−2.5	−2.5
Tryptophan	−3.4	−3.4

* Solvent parameter values assigned by Levitt (6).

Fig. 1, the hexapeptide analysis of sperm whale myoglobin, is illustrative. The high point of the profile, at position 60.5, falls within myoglobin antigenic site 2 (1). Several findings which proved to be generally true with other proteins can be seen in the myoglobin plot. First, not all antigenic determinants are associated with high points of hydrophilicity (for example, antigenic site 4, residues 113 through 119); second, not all high points are associated with antigenic determinants (position 79.5). The one correlation which has been upheld in myoglobin and the other proteins that we tested, is that one antigenic determinant is consistently located at the point of maximum hydrophilicity.

Computerization. To facilitate the analysis of large quantities of sequence information, our procedure was encoded in a FORTRAN program and run in a PDP 11/70 computer, and the resulting data was plotted with a Tektronix automatic plotting device.

List of Antigenic Determinants. Proteins with known antigenic determinants were considered to belong to one of two groups. Group 1, proteins whose antigenic structures are nearly or completely solved includes: (*i*) sperm whale myoglobin, with antigenic determinants at residues 15–22 (site 1), 56–62 (site 2), 94–99 (site 3), 113–119 (site 4), and 145–151 (site 5) (1); (*ii*) chicken lysozyme, with antigenic determinants including residues 5, 7, 13, 14, 33, 34, 62, 87, 89, 93, 96, 97, 113, 114, 116, and 125 (2); (*iii*) the ferredoxin from *Clostridium pasteurianum*, with antigenic determinants encompassing residues 1–7 and 51–55 (11); (*iv*) horse heart cytochrome *c*, with antigenic residues at positions 47, 58–62, 88–92, and 96 (4, 12); and (*v*) bovine myelin basic protein, with determinants in regions 64–73, 74–85, 113–121, and 153–166 (13, 14). Group 2, proteins for which partial information is available, comprises: (*i*) human hemoglobin β chains, with antigenic residues at 6, 16–23, 52, 68, 73, and 102 (3, 15, 16); (*ii*) the tobacco mosaic virus (vulgare) coat protein, with antigenic determinants at positions 62–68, 108–113, and 153–158 (17–19); (*iii*) human IgG heavy chain constant regions (each of the three constant domains of the Eu myeloma protein was considered as an individual protein), with antigenic determinants localized to position 214 of the CH1 domain, positions 296 and 309 of the CH2 domain, and 355 to 358 of the CH3 domain (20); (*iv*) bovine α-lactalbumin, where antigenic determinants have been located within residues 10–18, 60–80, 91–94, and 105–117 (unpublished data); and (*v*) leghemoglobin a from the soybean, with antigenic sites within residues 15–23, 52–59, 92–98, 107–116, and 132–142 (21).

Evaluating Predictions. An antigenic determinant was considered to be correctly identified by a prediction point if that point fell within the determinant, directly on a single antigenic residue, or within two residues (inclusive) on either side of any antigenic residue. This inclusion of a two residue "buffer zone" around antigenic sites is acceptable because much of the available information implicates single residues as antigenic sites, although in most cases these residues probably comprise part

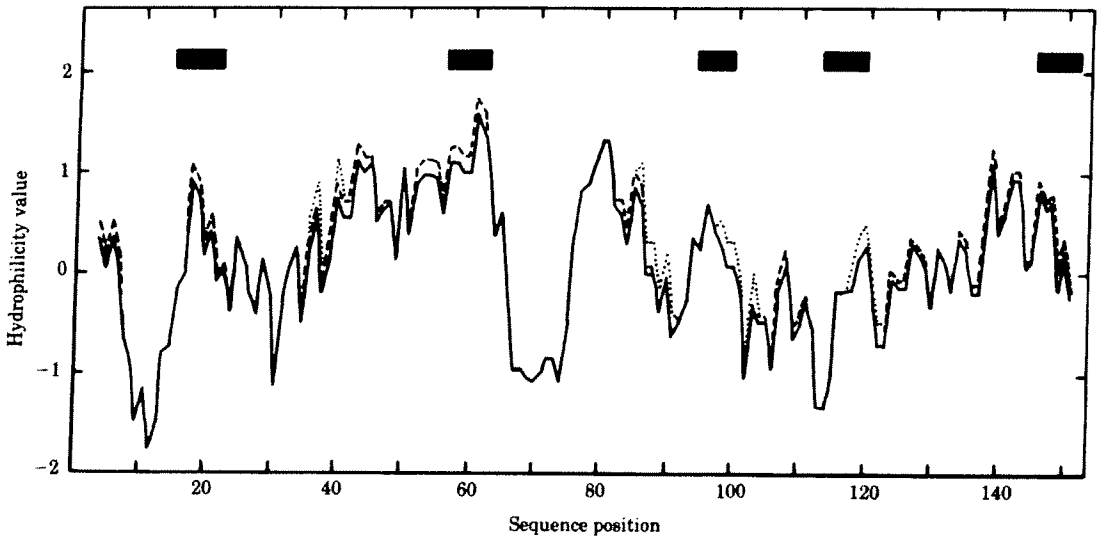


FIG. 1. Hexapeptide profile of sperm whale myoglobin. The averaged antigenicity values are plotted versus position along the amino acid sequence. The *x* axis contains 153 increments, each representing an amino acid in the sequence of myoglobin. The *y* axis represents the range of hydrophilicity values (from 3 to −3.4). The data points are plotted at the center of the averaging group from which they were derived. ■, Known antigenic determinants of myoglobin; ——, profile obtained by assigning the "solvent parameter" values of Levitt (6) to each amino acid; ----, profile obtained when the values for aspartic acid and glutamic acid were raised to 3.0; ·····, profile obtained when proline was assigned the value of 0.

Table 2. Prediction success with hexapeptide maximum point

	Correct	Wrong	Unknown
Original values	7	1	4
Asp, Glu = 3	8	0	4
Asp, Glu = 3, Pro = 0	10	0	2

of a larger site that includes several residues immediately adjacent to them in the sequence. Furthermore, any experiments designed to test the validity of these antigenic determinant predictions would be expected to include a number of residues on either side of the predicted point, in which case an overlap with the antigenic determinant would always be guaranteed.

Owing to the limited information available on the antigenic structures of some of the proteins used in this study, it was not always possible to definitely assess the correctness of a given prediction. Therefore, the proteins of groups 1 and 2 were treated differently in generating the information shown in Tables 2 and 3.

For group 1 proteins, a prediction was considered correct (Tables 2 and 3, column 1) if it successfully located an antigenic determinant or wrong (column 2) if it missed. With group 2 proteins, however, it is possible that a predicted point that misses *known* antigenic determinants may be indicating an antigenic determinant that is currently undiscovered. Therefore, for these proteins, predictions were considered to be correct (column 1) if they hit a known determinant or unknown if they missed (because they may yet prove to be hits).

Adjustment of Aspartic Acid, Glutamic Acid, and Proline Values. Table 2 shows the effect of increasing the values for these three amino acids from the original values given by Levitt. Increasing aspartic acid and glutamic acid from 2.5 to 3.0 eliminated the one wrong prediction and caused an elevation of the plots in many regions where antigenic determinants are known to exist (e.g., myoglobin sites 1, 2, and 5). There was no change in the number of unknown predicted points in the group 2 proteins (column 3), although the new values tended to elevate the profiles in the locations of the known determinants in these proteins. Next, the value for proline was raised to zero, and the hexapeptide analyses were repeated. The result is shown in line 3 of Table 2: two of the proteins that had given unknown predicted points now resulted in correct predictions.

The two remaining proteins with unknown prediction points are unusual, and it may not be worthwhile to attempt to bring them into the "correct" group by making further changes in amino acid values. For one of the two, the CH2 region of IgG, only 2 out of 109 residues are presently known to be antigenically active, and it may be possible that the predicted point may be indicating an undiscovered antigenic determinant. In the case of leghemoglobin a, the investigators indicate that they have not tested the antigenic activity of the predicted region (21).

Table 3. Effect of averaging group length on predictions by the maximum point

	Correct	Wrong	Unknown	C/C+W, %*
Dipeptide	23	17	18	58
Tripeptide	10	5	3	67
Tetrapeptide	9	3	4	75
Pentapeptide	8	2	2	80
Hexapeptide	10	0	2	100
Heptapeptide	7	3	2	70
Octapeptide	6	2	4	75
Nonapeptide	5	3	4	63
Decapeptide	5	2	5	71

* Percentage of correct assignments when considering only proteins of group 1. C, correct; W, wrong.

The Effect of Averaging Group Length. When only two amino acids are averaged at a time, the data plot is erratic, and the great variation in hydrophilicity over short lengths of peptide tends to obscure the general trend of the values. In addition, dipeptide analysis results in multiple identical high points because any pair of charged residues will yield the maximum value of 3.0. The results of this can be seen in line one of Table 3. For the 12 proteins analyzed, a total of 58 identical high points were obtained, and only 23 of these were associated with known antigenic determinants. Moreover, dipeptide analysis resulted in 17 wrong predictions.

Multiple identical high points continues to be a problem for tri- and tetrapeptide analysis; it finally disappears at the pentapeptide level (and higher), yielding a single predicted point for each of the 12 proteins. Although it is attractive to consider a method like the di-, tri-, or tetrapeptide analysis, which can predict more than one determinant per molecule, it seems more important to eliminate as many wrong predictions as possible because they reduce confidence in any given predicted antigenic determinant. As averaging group length increases, the number of wrong predictions decreases to a minimum of zero for hexapeptide analysis (Table 3). Comparison of data plots for various averaging group lengths suggested a reason for this. In going from di- to tetra- to hexapeptide analysis, the plots became less chaotic and the local hydrophilicity trend became more apparent. In going from hexa- to octa- to decapeptide analysis, the plots became even smoother. However, wrong predictions appeared again, and there was an increase in unknown predictions, whereas correct predictions fell from 10 to a low of 5 for nona- and decapeptide analysis. The reason for this may be that the regions of high hydrophilicity that are recognized well by the hexapeptide analysis begin to be obscured when longer averaging groups were used, due to their being combined with adjacent regions of low hydrophilicity.

Second and Third Highest Points. In order to assess the generality of the predictive value of high points, the success of the second and third highest points was considered. These points were only selected from the subset of points that had at least three amino acid positions between them and the highest (or second highest) point. This resulted in the second and third highest points always occurring in their own individual peak of hydrophilicity and the elimination of redundant prediction of antigenic determinants. However, neither the second nor the third highest points gave highly reliable prediction results. Although the correlation of predicted points with antigenic determinants seems to be significant in both cases (25% for the second and 33% for the third), the number of wrong predictions (33% in each case) severely limits the usefulness of these points for prediction of antigenic determinants of unknown proteins. These points are probably worthy of consideration in cases where immunochemical testing is used to verify the predictions because (by ignoring unknown predictions) they represent a 43% and 50% chance of a correct prediction, respectively.

Predictions for Uncharacterized Protein Antigens. We have applied our procedure to a number of proteins for which the location of an antigenic determinant may be of particular interest (Table 4). Several of the sequences listed in Table 4 are longer than six amino acids. In those cases, there are two or more adjacent sets of amino acids that result in identical average hydrophilicity values. Synthesis of short peptides should verify that these sequences are in, or immediately adjacent to, antigenic determinants.

To this end, we have recently used the Merrifield procedure to synthesize a peptide having the sequence αAbu-αAbu-Thr-

Table 4. Protein sequences with greatest average hydrophilicity*

Protein	Sequence
HBsAg (22)	141-Lys-Pro-Thr-Asp-Gly-Asn
Influenza hemagglutinins	
A/Victoria/3/75 strain (23)	171-Asn-Asp-Asn-Ser-Asp-Lys
A/Aichi/2/68 strain (24)	88-Val-Glu-Arg-Ser-Lys-Ala
Fowl plague virus hemagglutinin (25)	97-Glu-Arg-Arg-Glu-Gly-Asn
Human histocompatibility antigen HLA-B7 (26)	43-Pro-Arg-Glu-Glu-Pro-Arg
Human interferons	
Fibroblast (27)	103-Glu-Glu-Lys-Leu-Glu-Lys-Glu-Asp
Leukocyte I (28)	160-Glu-Arg-Leu-Arg-Arg-Lys-Glu
Leukocyte A (29)	131-Lys-Glu-Lys-Lys-Tyr-Ser
E. coli enterotoxins	
Heat labile (30)	66-Glu-Arg-Met-Lys-Asp-Thr
Heat stable (31)(two identical peaks)	26-Asp-Ser-Ser-Lys-Glu-Lys 46-Ser-Glu-Lys-Lys-Ser-Glu
Cholera toxin β chain (32)	79-Glu-Ala-Lys-Val-Glu-Lys
Streptococcal M protein (33)	58-Arg-Lys-Ala-Asp-Leu-Glu-Lys
Ragweed allergens	
Ra3 (34)	88-Cys-Thr-Lys-Asp-Gln-Lys
Ra5 (35)	40-Ser-Lys-Lys-Cys-Gly-Lys
Semliki Forest virus membrane proteins (36)	
E1	70-Thr-Lys-Glu-Lys-Pro-Asp
E2	246-Asp-Glu-Pro-Ala-Arg-Lys
E3	40-Glu-Asp-Asn-Val-Asp-Arg

* For each protein listed, the sequence of amino acids having the greatest average hydrophilicity value is shown; the number before the sequence indicates the position of the first amino acid in the group.

Lys-Pro-Thr-Asp-Gly-Asn-αAbu-Thr-αAbu (αAbu = α-amino butyric acid, replacing Cys) corresponding to residues 138–149 of the hepatitis B surface antigen (HBsAg) protein, and tested it for antigenic activity. The peptide side chains were deprotected under conditions where the peptide remained attached to the polystyrene beads (21). The peptidyl beads were then used to replace the polystyrene beads normally used in the Ausria II radioimmunoassay for HBsAg (Abbott), yielding a clearly positive binding affinity for ^{125}I-labeled anti-HbsAg antibodies. Beads without peptide, or peptidyl beads in which the side chain protecting groups had not been removed, did not bind significant ^{125}I-labeled anti-HBsAg antibody. Details of these experiments will be published elsewhere.

DISCUSSION

The studies described demonstrate the usefulness and limitations of antigenic determinant prediction by hydrophilicity analysis. The peak hexapeptide prediction value is highly successful, yielding no wrong assignments in 12 proteins; only lack of information on 2 of the 12 proteins makes it unclear whether the present method has a 100% success rate. On the other hand, the second and third highest peaks result in a mixture of correct and incorrect assignments and therefore, are less useful as predictors of antigenic determinants. It is clear by inspection of the data plots that some antigenic determinants are not correlated with hydrophilicity, although there does seem to be a correlation of many antigenic determinants with local upspikes of the hydrophilicity profile. This suggests that our present method may be a good basis on which to superimpose other types of information that boost the values of these low peaks. For example, it may be possible to improve prediction success by considering currently available methods for predicting secondary structure, particularly β bends.

Our method bears some resemblance to the procedure reported by Rose and Roy for predicting protein packing by hydrophobicity analysis (8), but it also has distinct differences that make it a better system for locating antigenic determinants. Because their approach utilizes the hydrophobicity values of Nozaki and Tanford (9) without the adjustments introduced by Levitt (6), the values for all hydrophilic amino acids are identical (i.e., 0), whereas the corresponding values used in our procedure range from 0.2 to 3.0. This results in a strong influence by the charged amino acids and an intermediate effect for neutral polar amino acids. Furthermore, Rose and Roy use a least-squares fitting of data to a quadratic polynomial with a seven-point moving window rather than hexapeptide averaging. This results in greater smoothing of the profile and end effects. Both of these qualities seem to decrease the potential usefulness for antigenic determinant prediction. In contrast, our method depends upon simpler calculations and a shorter averaging-group length and is capable of considering all amino acids from the amino-terminal to the carboxyl-terminal residue.

Finally, it should be emphasized that the ability to predict antigenic determinants from amino acid sequence data alone is potentially very useful, even though only a single determinant can be predicted with confidence for any given molecule. For example, many proteins whose antigenic structures are of interest are not available in quantities sufficient to allow conventional immunochemical studies to be carried out, as is the case with many of the proteins for which we listed predictions in the preceding section. Increasingly, amino acid sequence information for such proteins is being obtained by microchemical methods or by nucleotide sequence analysis, so that sufficient material for conventional immunochemical analysis is never available. However, once an antigenic determinant has been predicted, it should be possible to verify its existence by synthesizing the indicated region chemically and testing its activity in an appropriate immune assay, such as inhibition of cytotoxicity or precipitation inhibition. Furthermore, it should be possible to raise antisera against such synthetic determinants, as Arnon *et al.* have done for a bacteriophage (37). Ultimately, predicted antigenic determinants from proteins of pathogenic organisms might be useful in the production of synthetic vaccines.

1. Atassi, M. Z. (1975) *Immunochemistry* **12**, 423–438.
2. Atassi, M. Z. & Lee, C. L. (1978) *Biochem. J.* **171**, 429–434.
3. Reichlin, M. (1975) *Adv. Immunol.* **20**, 71–123.
4. Jemmerson, R. & Margoliash, E. (1979) *J. Biol. Chem.* **254**, 12706–12716.
5. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–224.
6. Levitt, M. (1976) *J. Mol. Biol.* **104**, 59–107.
7. Maxfield, F. R. & Scheraga, H. A. (1976) *Biochemistry* **15**, 5138–5153.
8. Rose, G. D. & Roy, S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4643–4647.
9. Nozaki, Y. & Tanford, C. (1971) *J. Biol. Chem.* **246**, 2211–2217.
10. Kabat, E. A. (1968) *Structural Concepts in Immunology and Immunochemistry*, (Holt, Rinehart & Winston, New York), pp. 89–100.
11. Kelly, B. & Levy, J. G. (1971) *Biochemistry* **10**, 1763.
12. Jemmerson R. & Margoliash, E. (1977) *Adv. Exp. Med. Biol.* **98**, 119.
13. Eylar, E. H. (1977) *Adv. Exp. Med. Biol.* **98**, 259.
14. Westall, F. C. & Thompson, M. (1978) *Immunochemistry* **15**, 189.
15. Kazim, A. L. & Atassi, M. Z. (1977) *Biochem. J.* **167**, 275.
16. Reichlin, M. & Noble, R. W. (1977) in *Immunochemistry of Proteins*, ed. Atassi, M. Z. (Plenum, New York), Vol. 2, p. 311.

17. Anderer, F. A. (1963) Z. *Naturforsch.* **18b**, 1010.
18. Benjamini, E. (1977) in *Immunochemistry of Proteins*, ed. Atassi, M. Z. (Plenum, New York), Vol. 2, p. 265.
19. Milton, R. C. Del. & van Regenmortel, M. H. V. (1979) *Mol. Immunol.* **16**, 179.
20. Kehoe, M. J. & Seide-Kehoe, R. (1979) in *Immunochemistry of Proteins*, ed. Atassi, M. Z. (Plenum, New York), Vol. 3, p. 87.
21. Hurrell, J. G. R., Smith, J. A. & Leach, S. J. (1978) *Immunochemistry* **15**, 297–302.
22. Valenzuela, P., Gray, P., Quiroga, M., Zaldivar, J., Goodman, H. M. & Rutter, W. J. (1979) *Nature (London)* **280**, 815–818.
23. Jou, W. M. Verhoeyen, M., Devos, R., Saman, E., Fang, R., Huylebroeck, D., Fiers, W., Threlfall, G., Barber, C. & Carey, N. (1980) *Cell* **19**, 683–696.
24. Verhoeyen, M., Fang, R., Jou, W. M., Devos, R., Huylebroeck, D., Saman, E. & Fiers, W. (1980) *Nature (London)* **286**, 771–776.
25. Porter, A. G., Barber, C., Carey, N. H., Hallewell, R. A., Threlfall, G. & Emtage, J. S. (1979) *Nature (London)* **282**, 471–477.
26. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. L. (1979) *Biochemistry* **18**, 5711–5720.
27. Derynck, R., Content, J., DeClercq, E., Volckaert, G., Tavernier, J., Devos, R. & Fiers, W. (1980) *Nature (London)* **285**, 542–547.
28. Taniguchi, T., Mantei, N., Schwarzstein, M., Nagata, S., Muramatsu, M. & Weissman, C. (1980) *Nature (London)* **285**, 547–549.
29. Goeddel, D. V., Yelverton, E., Ullrich A., Heyneker, H. L., Miozzari, G., Holmes, W., Seeburg, P. H., Dull, T., May, L., Stebbing, N., Crea, R., Maeda, S., McCandliss, R., Sloma, A., Tabor, J. M., Gross, M., Familletti, P. C. & Pestka, S. (1980) *Nature (London)* **287**, 411–416.
30. Dallas, W. S. & Falkow, S. (1980) *Nature (London)* **288**, 499–501.
31. So, M. & McCarthy, B. J. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4011–4015.
32. Kurosky, A., Markel, D. E., Peterson, J. W. & Fitch, W. M. (1977) *Science* **195**, 299–301.
33. Beachey, E. H., Seyer, J. M. & Kang, A. H. (1980) *J. Biol. Chem.* **255**, 6284–6289.
34. Klapper, D. G., Goodfriend, L. & Capra, J. D. (1980) *Biochemistry* **19**, 5729–5734.
35. Mole, L. E., Goodfriend, L., Lapkoff, C. B., Keoho, J. M. & Capra, J. D. (1975) *Biochemistry* **14**, 1216–1220.
36. Caroff, H., Frischauf, A. M., Simons, K., Lehrach, H. & Delius, H. (1980) *Nature (London)* **288**, 236–241.
37. Arnon, R., Sela, M., Parant, M. & Chedid, L. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6769–6772.

EXHIBIT 6

☐ **1:** AAA72260. Reports human growth horm...[gi:208528]

BLink, Conserved Domains, Links

Comment Features Sequence

```
LOCUS       AAA72260                 192 aa            linear   SYN 27-APR-1993
DEFINITION  human growth hormone.
ACCESSION   AAA72260
VERSION     AAA72260.1  GI:208528
DBSOURCE    locus SYNHUMGHS accession K02382.1
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (residues 1 to 192)
  AUTHORS   Ikehara,M., Ohtsuka,E., Tokunaga,T., Taniyama,Y.O., Iwai,S.,
            Kitano,K., Miyamoto,S., Ohgi,T., Sakuragawa,Y., Fujiyama,K.,
            Ikari,T., Kobayashi,M., Miyake,T., Shibahara,S., Ono,A., Ueda,T.,
            Tanaka,T., Baba,H., Miki,T., Sakurai,A., Oishi,T., Chisaka,O. and
            Matsubara,K.
  TITLE     Synthesis of a gene for human growth hormone and its expression in
            Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 81 (19), 5956-5960 (1984)
   PUBMED   6091124
COMMENT     [1] synthesized this gene using the phosphotriester method with
            frequently occuring amino acid codons of E. coli.  When the gene
            was inserted into an E. coli plasmid used to transform E.  coli
            cells, a polypeptide identical to natural human growth hormone was
            produced.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..192
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
     Protein         1..192
                     /name="human growth hormone"
     Region          5..190
                     /region_name="Hormone_1"
                     /note="Somatotropin hormone family; pfam00103"
```

```
                     /db_xref="CDD:109170"
     CDS             1..192
                     /coded_by="K02382.1:5..583"
                     /transl_table=11
ORIGIN
        1 mfptiplsrl fdnamlrahr lhqlafdtyq efeeayipke qkysflqnpq tslcfsesip
       61 tpsnreetqq ksnlellris llliqswlep vqflrsvfan slvygasdsn vydllkdlee
      121 giqtlmgrle dgsprtgqif kqtyskfdtn shnddallkn ygllycfrkd mdkvetflri
      181 vqcrsvegsc gf
//
```